

Enhancing low-light pedestrian detection: convolutional neural network and YOLOv8 integration with automated dataset

Rendi, Devi Fitriana

Department of Computer Science, Binus Graduate Program, Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia

Article Info

Article history:

Received Jun 23, 2024

Revised Dec 14, 2024

Accepted Dec 25, 2024

Keywords:

Computer vision

Convolutional neural network

Deep learning

Low light enhancement

Pedestrian detection

ABSTRACT

This research aims to enhance the you only look once (YOLO) model for pedestrian detection in environments with varying lighting conditions, particularly in low-light scenarios. The primary contribution of this work is the integration of a convolutional neural network (CNN)-based low-light enhancement model, which transforms dark images into brighter, more discernible ones. This enhanced dataset is subsequently used to train the YOLO model, allowing it to learn from both the original and transformed data distributions. Unlike traditional YOLO training approaches, this method generates more accurate data representations in challenging lighting environments, leading to improved detection outcomes. The novelty of this approach lies in its dual-stage training process, which integrates a CNN-based low-light enhancement model with YOLO's detection capabilities. This combination not only enhances pedestrian detection but also has the potential for application in other domains, such as vehicle detection and surveillance, particularly in challenging lighting conditions. The automatic dataset collection pipeline provides an efficient way to gather diverse training data across various scenarios. The YOLOv8 model trained on the low-light enhanced dataset significantly outperformed the baseline model trained only on the original dataset, with precision increased by 9.8%, recall by 45.7%, mAP50 by 26.8%, and mAP50-95 by 41.0% when validated on dark images.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Rendi

Department of Computer Science, Binus Graduate Program, Master of Computer Science

Bina Nusantara University

Jakarta, Indonesia

Email: rendi002@binus.ac.id

1. INTRODUCTION

Pedestrian detection systems are integral to autonomous driving, providing critical safety and navigation capabilities. Recent studies, such as "Localized semantic feature mixers for efficient pedestrian detection in autonomous driving" [1], have demonstrated promising results across various global datasets. Additionally, the research titled "Faster region-based convolutional neural network (R-CNN) deep learning model for pedestrian detection from drone images" [2] highlighted the effectiveness of you only look once (YOLO) and faster R-CNN algorithms in achieving high accuracy levels in drone imagery. However, despite these advancements, practical implementations of pedestrian detection systems encounter significant challenges. The diverse nature of environments, particularly variations in lighting conditions, directly impacts the accuracy and reliability of these systems.

The YOLO model has emerged as one of the leading approaches in real-time object detection, demonstrating remarkable speed and accuracy in detecting pedestrians. Nevertheless, achieving high

accuracy requires large, diverse training datasets, often challenging to obtain. This is where the concept of streaming data-generated datasets becomes particularly relevant. By employing streaming data collection methods, we can create larger and more representative datasets, enhancing the performance of pedestrian detection models. However, these streaming datasets introduce complexities, including intricate data processing and a deep understanding of the factors influencing pedestrian detection.

This paper aims to address these challenges by integrating a CNN-based low light enhancement model into pedestrian detection systems, focusing on enhancing performance in suboptimal lighting conditions where traditional algorithms may struggle. The novelty of our dual-stage training process significantly enhances the YOLO model's ability to detect pedestrians in varying lighting environments, resulting in substantial performance improvements compared to baseline models.

Additionally, we propose a framework for streaming dataset collection to complement the integration of the low light enhancement model. This framework incorporates real-time data acquisition methods from multiple sources, such as cameras and sensors, facilitating the continuous generation of a diverse and extensive dataset. By utilizing streaming data, we ensure the availability of up-to-date and representative data for training and testing pedestrian detection models, significantly improving their robustness and adaptability in real-world scenarios.

In summary, our contributions can be delineated as follows: i) we propose a training mechanism that integrates outputs from the low light enhancement model into the training process of pedestrian detection systems, allowing the system to learn the data distribution derived from enhanced images. This integration aims to improve the overall detection performance quantitatively; ii) we introduce a comprehensive framework for streaming dataset collection that enhances the integration of the low light enhancement model; and iii) we conduct an examination of various factors influencing pedestrian detection performance, including environmental conditions, providing valuable insights for future advancements in pedestrian detection technology.

2. RELATED WORKS

The YOLOv8 approach has proven effective in rapid object detection, while the low light enhancement CNN model has been identified as a potential solution to address the common challenges of low-light conditions encountered in CCTV recordings. Studies [1], [2] highlight specific pedestrian detection approaches, along with findings from [3], [4] regarding YOLO and YOLOv8, provide a solid foundation for building a pedestrian detection model.

Recent research emphasizes the potential of integrating advanced techniques such as domain adaptation, transfer learning, and generative methods to improve object detection performance across various environments [5]-[9]. These findings suggest that similar approaches may enhance pedestrian detection capabilities, particularly in suboptimal lighting conditions when integrated with the low light enhancement model and YOLOv8.

Further exploration into deep learning techniques reveals their capacity to enhance visual information in challenging scenarios. For instance, Kuang *et al.* [10] leverages a conditional generative adversarial network (CGAN) to transform grayscale thermal images into realistic RGB counterparts, while Qiu *et al.* [11] proposes IDOD-YOLOv7, a method combining dehazing and a self-adaptive image processing module with a pre-trained YOLOv7 network to address object detection in low-light, foggy traffic environments. These investigations underscore the potential of deep learning for enriching visual data across diverse domains.

Additionally, studies [12]-[19] propose training deep neural networks to process low-light images through techniques like color transformations, noise reduction, histogram equalization, and self-calibrated illumination frameworks. These conclusions support the integration of the low light enhancement model to improve the quality of data used in pedestrian detection, thereby adding significant value to this research. The authors in [20] collected a dataset, creating normal and low-light image twins known as the LOL dataset, simulating normal lighting conditions for corresponding low-light images through specific transformations.

Moreover, Cauldron [21] discusses modifications made to the YOLOv3 object detection model, enhancing its performance for vision-impaired users in low-light environments. Rather than directly training YOLO on dark images, the author proposes utilizing a learned low-light image processing model called *see in the dark* (SID) as a pre-processing pipeline before running YOLO. Similarly, Roy and Bhaduri [22] explores the DenseSPH-YOLOv5 model, which combines DenseNet blocks with a swin-transformer prediction head, showcasing the integration of advanced architectures to improve object detection in complex scenarios. Research by Roy *et al.* [23] also introduce a fast and accurate fine-grain object detection model based on YOLOv4, highlighting its versatility across applications, including plant disease detection.

This literature review provides a strong theoretical basis for combining the low light enhancement CNN model with YOLOv8 in pedestrian detection. By leveraging the strengths of each model and responding to recent findings in the literature, this research is expected to contribute to the development of efficient and reliable solutions for pedestrian detection in CCTV recording environments with varying lighting conditions.

3. METHOD

3.1. Framework for streaming dataset generation

Camera placement has been strategically placed out to collect a dataset specifically focusing on pedestrian behavior. These cameras are positioned at optimal heights and strategic locations, providing broad coverage of significant pedestrian traffic areas. Particularly for facial identification and other important attributes. Avoiding placement that is too high, which may reduce detail, or too low, which may obstruct the view.

This research adjusts the focused area coverage to include the most frequently used pedestrian pathways while avoiding recording areas that may involve individuals' privacy, thus complying with applicable ethical standards and privacy regulations.

As shown in Figure 1, the dataset preparation process for the YOLOv8 model in pedestrian detection from CCTV footage involves several crucial steps. Firstly, after recording the video from CCTV, data can be sent to the system using an application programming interface (API). It is then processed to extract individual frames at a 3-second interval. This approach provides a sufficiently good representation of pedestrian movements and positions over a certain period.

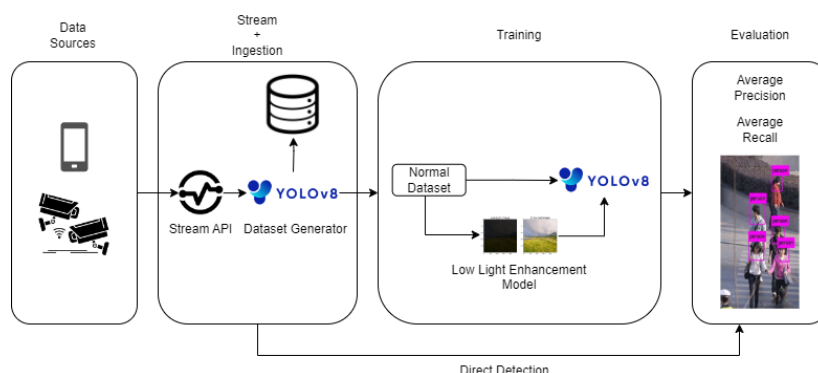


Figure 1. Streaming dataset collection framework

After obtaining this set of frames, the next step is to utilize the pre-trained YOLOv8 model trained on the COCO dataset [24]. The pretrained model provides a broad initial knowledge of various objects in a general context, including humans (people). Therefore, the model can better capture features specific to pedestrian detection.

Subsequently, the inference process is conducted using the YOLOv8 model on each frame, with a specific focus on the "people" class. By limiting attention to this class, the model concentrates on identifying and tagging pedestrian locations in each frame. This process enables the automatic creation of bounding boxes around detected human objects, providing the necessary location and size information for the next stage, annotation.

However, in some intervals of the footage, no pedestrians were present, so those images need to be removed from the dataset to maintain relevance. The elimination process relied on the YOLOv8 model for label detection. If the model failed to detect any label in each image, indicating the absence of pedestrians, the image was flagged for deletion from the dataset. This ensured that only images containing pedestrian labels were used for further processing and analysis.

Annotation is a crucial step in dataset development, and by using the inference results from the YOLOv8 model, the generated bounding boxes can be considered as automatic annotations. Each bounding box, signifies the location of pedestrians in that frame. This process significantly accelerates dataset creation, reducing the manual effort required to label each object individually.

<object-class> is an integer representing the object class. Class indices must start from 0 and increase by 1 for each unique class in the dataset. <x-center> and <y-center> are the coordinates of the

bounding box center, each normalized based on the width and height of the image. Values should be in the range of 0 to 1. <width> and <height> are the width and height of the bounding box, each normalized based on the width and height of the image. Values should be in the range of 0 to 1.

The dataset is split into three sections: the training set (60%), used to teach the model; the validation set (30%), used to check how well the model learns; and the test set (10%), used to see how good the model is after learning. Afterwards, these sets are organized into folder structures.

3.2. Low light image enhancement model

Advanced processing techniques, such as scaling or histogram stretching, can be applied, but they do not address the low signal-to-noise ratio (SNR) caused by the low number of photons. The camera is not able to take the perfect image at low light because of the noise created at the camera sensors. During night mode the noise is cancelled out by the camera to get the perfect image [12]. Various denoising, deblurring, and enhancement techniques like [11] have been proposed, but their effectiveness is limited in extreme conditions, such as nighttime imaging at high speeds. However, solely increasing the brightness of dark regions will inevitably amplify image degradation.

There are physical ways to improve SNR in low-light conditions, including widening the aperture, lengthening exposure time, and using flash as stated in [17]. However, each has its own characteristic limitations. For example, lengthening exposure time can result in blur due to camera shake or object movement.

To address the challenges posed by low-light conditions, a promising solution is the utilization of a low light enhancement model. The model for low light enhancement that we are utilizing is inspired by [16]. This approach offers a practical means to mitigate the limitations associated with traditional methods like widening the aperture or lengthening exposure time. It often helps to visualize its internal structure. One way to do this is by displaying images from each layer of the model, as shown in Figure 2.

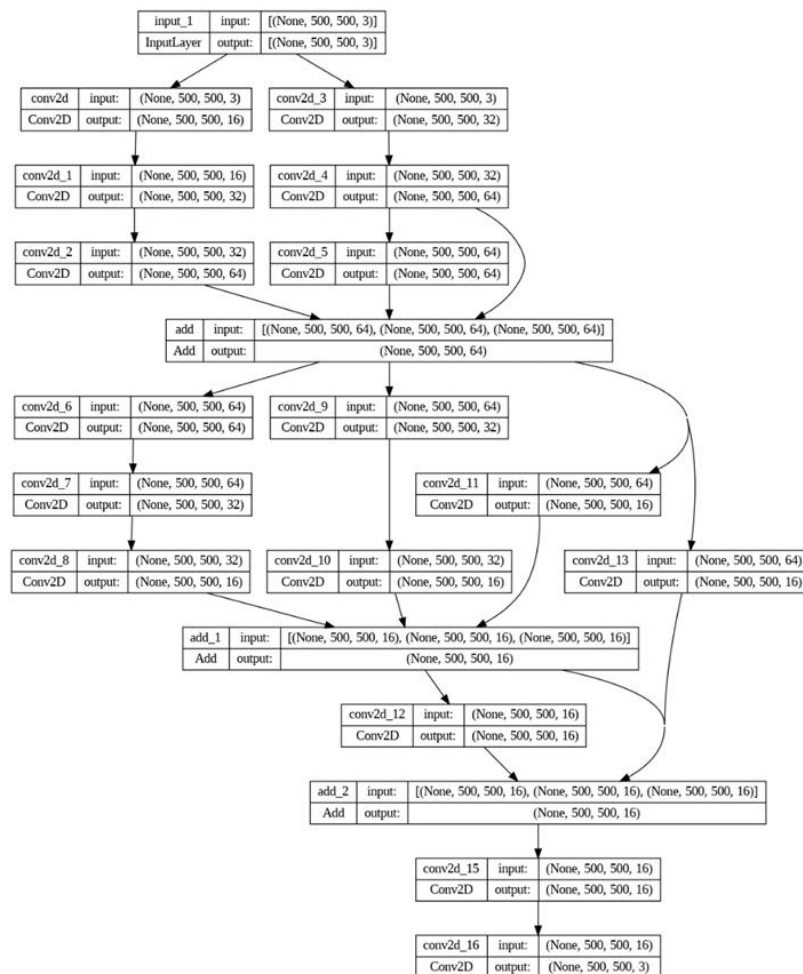


Figure 2. Low light enhancement model visualization

While some research explores networks that require specialized data formats, following the path of [19], this model offers a practical advantage by working with standard image formats like JPEG and PNG making it more suitable for real-world applications.

A resizing process is implemented following the application of the low light enhancement model to address the dimensional changes produced by the model. After enhancing the brightness of the images, the dimensions of the processed images are reverted to their original size before being input into the YOLOv8 model. This step is crucial for maintaining consistent image dimensions, significantly contributing to the accuracy and efficiency of pedestrian detection, particularly in suboptimal lighting conditions. This adaptation serves as a key element in the workflow of this study, ensuring a seamless integration between image quality enhancement and the object detection process.

Supervised learning is a machine learning approach that's defined by its use of labeled datasets [25], meaning each input data point is associated with a corresponding target or output. In the context of the low light enhancement (CNN), each input image in the training dataset must have a corresponding ground truth enhanced image to guide the learning process effectively. By incorporating simulation algorithms to augment the dataset, the model is exposed to a wider range of low-light scenarios, enhancing its ability to generalize and produce high-quality enhancements in various real-world conditions.

3.3. Low light image augmentation

Training a supervised low light enhancement model typically requires paired low-light and normal-light images. Various approaches have been used in previous research to collect the necessary dataset. Research [12] involves using collected a dataset of short exposure. While [16] simulates low-light image condition using algorithms like salt and pepper and gamma reduction are applied to the dataset.

It is important to note that the dataset used in training the low light enhancement model does not have paired lighting conditions needed for training. To address this limitation, this study decided to introduce variations in low-light image conditions by incorporating several simulation algorithms. To simulate low-light image condition, algorithms like salt and pepper and gamma reduction are applied to the dataset, as shown in Figure 3. The salt and pepper algorithm are used to introduce random noise into the pixels of the image, while the gamma reduction algorithm is used to reduce the brightness of the image. Figure 3(a) shows the original image without any modifications, Figure 3(b) illustrates the image with added salt and pepper noise, Figure 3(c) represent the image after applying both salt and pepper and gamma reduction. By introducing these variations, the CNN model can be trained to recognize and enhance images under suboptimal lighting conditions, even though the main dataset does not encompass a variety of lighting conditions.



Figure 3. Image processing stages; (a) original image, (b) image with salt-and-pepper noise, and (c) image with salt-and-pepper noise and gamma reduction

Let I be the input image with dimensions $\text{row} \times \text{col} \times \text{ch}$. Let $S_{vs,p}$ represent the salt vs. pepper ratio, and amount denote the noise amount. The number of salt pixels (num_{salt}) and pepper pixels ($\text{num}_{\text{pepper}}$) are calculated based on the noise amount and salt vs. pepper ratio as in (1) and (2). The coordinates of salt and pepper pixels are randomly selected within the image dimensions. Finally, the salt and pepper pixels are set to their respective intensity values. This equation represents the salt and pepper noise generation process in the code.

$$\text{num}_{\text{salt}} = [\text{amount} \times \text{size}(I) \times S_{vs,p}] \quad (1)$$

$$\text{num}_{\text{pepper}} = [\text{amount} \times \text{size}(I) \times 1 - S_{vs,p}] \quad (2)$$

To analyze both unprocessed and processed images, we used grayscale histograms to assess their similarity. “Histograms in general are frequency distributions, and histograms of images describe the frequency of the intensity values that occur in an image” [26].

3.4. Training YOLOv8 using enhanced low light images

In this study, propose an approach involving the use of enhanced image quality results from the low light enhancement model as training dataset for the YOLOv8 model. This approach aims for the model to learn the changes in data distribution that occur in images processed through low light enhancement model.

During the training process of YOLO for pedestrian detection, images that have been processed by a low light enhancement model are utilized. In this scenario, YOLOv8 is employed with various model sizes, namely YOLOv8n, YOLOv8m, and YOLOv8x [27] for comparison purposes. This approach aims to assess the performance and efficiency of different YOLOv8 configurations in detecting pedestrians in low-light conditions, thereby informing decisions on model selection and optimization for improved accuracy and reliability.

For this research, the default configuration was used for training YOLOv8. For each YOLOv8 variant, training was conducted with varying numbers of epochs 10, 20, 30, 40, and 50 as base standards, learning rate of 0.002, image size of 640, and batch size 16. The purpose of training these models with different numbers of epochs is to evaluate their performance and convergence over time.

3.5. Evaluation criteria

Evaluation criteria for the low-light enhancement model involve comparing histograms of the original and enhanced images. First, each image is converted from red, green, blue (GRB) to hue, saturation, value (HSV) color space, which helps in separating hue from intensity, facilitating better analysis of brightness and color consistency. Histograms are then computed for each HSV channel to represent pixel intensity distributions, and each histogram is normalized for consistency. These normalized histograms are merged into a feature vector to capture the image’s overall color and brightness distribution. To evaluate similarity, a correlation method is used to compare the feature vectors of the original and enhanced images, providing a similarity score. The average similarity score is then calculated across all image pairs, using (3):

$$\text{Average Score} = \frac{1}{N} \sum_{i=0}^n C(\hat{H}_{\text{original},i}, \hat{H}_{\text{enhanced},i}) \quad (3)$$

where $C(\hat{H}_{\text{original},i}, \hat{H}_{\text{enhanced},i})$ is the correlation score for each original-enhanced image pair. This average score reflects the model’s overall effectiveness in enhancing low-light images while retaining natural image quality.

The main criteria used for target detection are mainly precision, recall, mean average precision (mAP), and other relevant measures. Precision evaluates the ratio of correctly identified objects to the total number of detections, as in (4). Recall measures its capability to detect all instances of objects present in the images, as in (5). Intersection over union (IoU) measures the degree of overlap between a predicted bounding box and an actual bounding box. It plays a fundamental role in evaluating the accuracy of object localization as in (6).

$$\text{Precision} = \frac{TP}{TP+FP} \quad (4)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (5)$$

$$\text{IoU} = \frac{\text{Area of Intersection of two boxes}}{\text{Area of Union of two boxes}} \quad (6)$$

mAP50 is a mean average precision at an IoU threshold of 0.50 offers a specific measure of the model’s accuracy, particularly for objects that are easily detectable. While mAP50-95 is mean average precision calculated across IoU thresholds between 0.50 and 0.95. It gives a comprehensive view of the model’s performance across different levels of detection difficulty.

4. RESULTS AND DISCUSSION

4.1. Generated dataset

The framework proposed tested on video data with duration around 4 to 5 hours managed to extract a total of 4,994 images from the video. It’s important to note that this dataset primarily consists of indoor

photos taken within a shopping. These images were captured using the Xiaomi C300 CCTV with a 2K resolution and F1.4 aperture.

Following the removal process of images lacking pedestrian labels, the dataset was reduced to 4,970 images. These images were then divided into three folders for training, validation, and testing purposes. The ratio between the three sets was as follows: the training set contained approximately 60% of the images (2,982 images), the validation set contained approximately 30% (1,491 images), and the testing set contained the remaining approximately 10% (497 images) of the dataset.

After the data split, each folder was supplemented with a data.yaml file specifying the dataset for YOLOv8. The file contains information about the location of images and labels for each part of the dataset, namely training, validation, and testing. Each part has a path entry for the location of images and labels for the location of corresponding labels. There is also a nc entry indicating the number of classes in the dataset, and names which is a list of class names which is set to 1 since we are detecting only one class, there is only one class label, which is "people".

4.2. Low light enhancement model

The training process progressed smoothly and consistently, with each epoch contributing to the model's refinement. To evaluate the performance of the low-light enhancement model, various metrics were analyzed to assess its effectiveness in improving image quality under low-light conditions. These metrics provided quantitative insights into the model's capabilities, offering a systematic assessment of its performance, as illustrated in Figure 4.

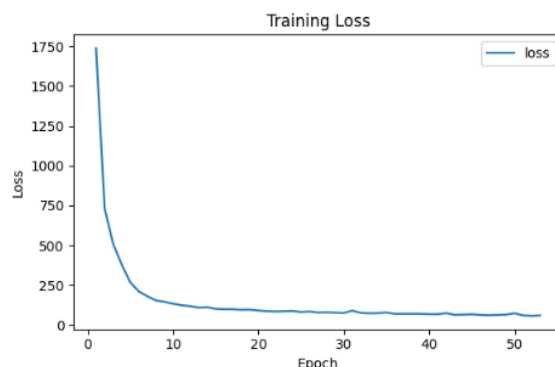


Figure 4. Low light enhancement model training loss

In addition, a quantitative evaluation was conducted to assess the similarity between the original images and those processed with the low-light enhancement model, specifically within the test dataset. The method used was histogram comparison, where each image was converted into the HSV color space to produce a more stable histogram representation. Histograms were calculated for each channel and normalized. The histograms of the original and enhanced images were then compared using the correlation method, which measures the degree of similarity between the two histogram distributions. The processed images achieve a strong similarity to the originals, suggesting the processing preserves the image integrity very well, as shown in Figure 5. Figure 5(a) shows the histogram of the original image, Figure 5(b) illustrates the histogram of the dark image after undergoing low-light enhancement.

This process was applied to all image pairs in the test dataset, where images from both the original and enhanced folders were matched and compared individually. After completing all comparisons, the average histogram comparison score was calculated, yielding a result of 0.836. This indicates a high degree of similarity between the original and enhanced images, suggesting that the enhanced images retained visual characteristics closely resembling the original ones, despite undergoing low-light enhancement.

While quantitative metrics provide valuable numerical insights, they may not fully capture the qualitative aspects of image enhancement. A qualitative assessment, which involves visually inspecting the enhanced images and evaluating their perceptual quality, offers a more comprehensive understanding of the model's capabilities. By combining both quantitative metrics and qualitative evaluations, we can ensure that the low-light enhancement model's performance is thoroughly assessed based on its ability to produce visually pleasing and perceptually enhanced images, as shown in Figure 6. Figure 6(a) shows the original or ground truth image, Figure 6(b) presents the low-light image before enhancement, and Figure 6(c) displays the enhanced image after applying the low-light enhancement model.

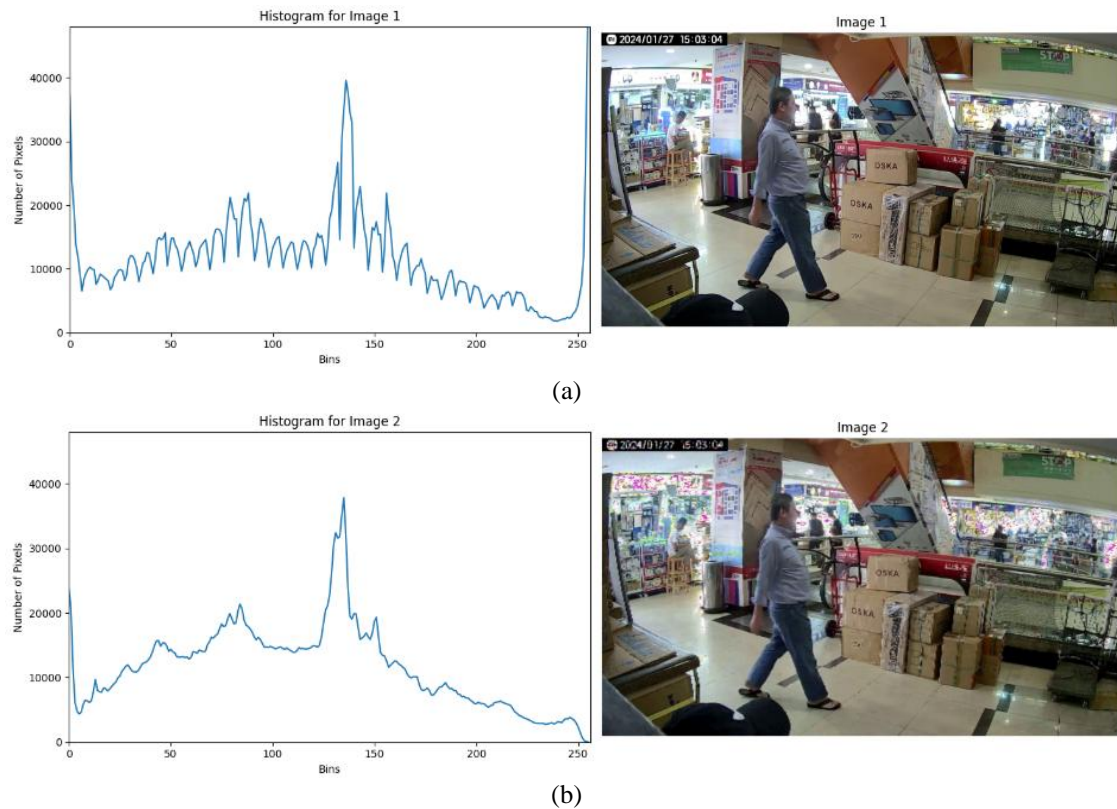


Figure 5. Histogram comparison; (a) original histogram and (b) processed grayscale histogram

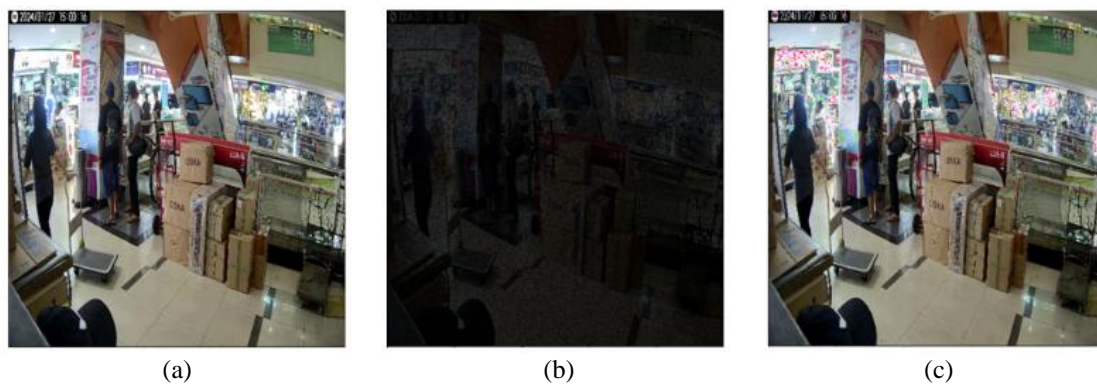


Figure 6. Comparison of image enhancement stages; (a) original or ground truth image, (b) low-light image, and (c) enhanced image after applying the low-light enhancement model

4.3. Pedestrian detection YOLOv8 model

Table 1 presents the validation results for precision, recall, mAP50, and mAP50-95 obtained from three variants of YOLOv8: YOLOv8n, YOLOv8m, and YOLOv8x. These models were trained using both the original dataset and the dataset processed by the low-light enhancement model.

The validation was conducted exclusively on images pre-processed through the low-light enhancement model, under the assumption that all images would undergo this preprocessing step before detection. As shown in Table 1, the training consistently yielded higher mAP values for models trained using the enhanced dataset compared to those trained on the original dataset.

Figure 7 shows a batch image from the YOLOv8x validation results, demonstrating the model's accuracy in detecting objects when using the CNN-processed dataset. The bounding boxes in the figure indicate that the YOLOv8x model correctly identifies individuals in the image.

Table 1. YOLOv8 original and enhanced dataset validation result

No	Description	Results				
		Epoch	Precision	Recall	mAP50	mAP50-95
1	YOLOv8n trained with original dataset	10	0.745	0.711	0.775	0.484
		20	0.821	0.751	0.835	0.591
		30	0.841	0.767	0.853	0.625
		40	0.854	0.768	0.858	0.642
		50	0.85	0.781	0.861	0.653
2	YOLOv8n trained with CNN processed dataset	10	0.782	0.684	0.776	0.493
		20	0.824	0.759	0.842	0.59
		30	0.851	0.763	0.855	0.622
		40	0.846	0.78	0.863	0.642
		50	0.851	0.787	0.865	0.657
3	YOLOv8m trained with original dataset	10	0.816	0.764	0.838	0.601
		20	0.858	0.787	0.869	0.677
		30	0.852	0.813	0.873	0.7
		40	0.859	0.804	0.877	0.71
		50	0.875	0.795	0.879	0.72
4	YOLOv8m trained with CNN processed dataset	10	0.828	0.764	0.842	0.608
		20	0.859	0.794	0.878	0.682
		30	0.874	0.802	0.885	0.711
		40	0.859	0.819	0.888	0.722
		50	0.874	0.81	0.891	0.732
5	YOLOv8x trained with original dataset	10	0.839	0.764	0.846	0.628
		20	0.865	0.788	0.876	0.699
		30	0.861	0.806	0.879	0.722
		40	0.865	0.81	0.883	0.731
		50	0.877	0.806	0.882	0.736
6	YOLOv8x trained with CNN processed dataset	10	0.827	0.775	0.849	0.623
		20	0.863	0.805	0.884	0.704
		30	0.864	0.818	0.889	0.732
		40	0.868	0.817	0.894	0.742
		50	0.878	0.816	0.894	0.75



Figure 7. YOLOv8x trained with CNN processed dataset validation prediction

In addition to validation with the low-light enhanced dataset, validation was also performed on images that underwent augmentation (darkening). The evaluation of YOLOv8's object detection performance showed significant differences between the two approaches. YOLOv8x-CNN, trained using the low-light enhanced dataset, demonstrated superior performance at epoch 50 with a precision of 0.878, recall of 0.816, mAP50 of 0.894, and mAP50-95 of 0.75. In contrast, YOLOv8x-Orig-Dark, trained on the original dataset and validated on darkened images, performed worse with a precision of 0.800, recall of 0.560, mAP50 of 0.705, and mAP50-95 of 0.532. Table 2 shows the results of this validation, providing evidence that the low-light enhancement model offers superior performance compared to darkened images processed via augmentation alone.

The high precision observed in the darkened dataset indicates that the model maintains accuracy in detecting recognized objects. However, the lower recall reveals difficulty in identifying all objects in darker images. This precision-recall imbalance suggests that while the model ensures correct detections, it often misses objects in low-light scenarios. Figure 8 shows validation results on darkened images, where the YOLOv8 model trained on the original dataset struggled to produce accurate bounding boxes on augmented dark images.

Table 2. YOLOv8 dark dataset validation result

No	Description	Epoch	Results			
			P	R	mAP50	mAP50-95
1	YOLOv8x-Orig-Dark	10	0.510	0.129	0.317	0.206
		20	0.804	0.272	0.545	0.382
		30	0.878	0.388	0.643	0.469
		40	0.832	0.407	0.633	0.469
		50	0.800	0.560	0.705	0.532
2	YOLOv8x-CNN-Dark	10	0.731	0.175	0.455	0.285
		20	0.803	0.488	0.666	0.464
		30	0.890	0.511	0.715	0.533
		40	0.719	0.556	0.672	0.495
		50	0.895	0.590	0.761	0.586

*Orig: trained with original dataset
*CNN: trained with processed dataset
*Dark: validate dark dataset



Figure 8. YOLOv8x validation on dark dataset prediction

5. CONCLUSION

In conclusion, this study demonstrates significant improvements in pedestrian detection performance by integrating the YOLOv8 model with a low light enhancement CNN. The YOLOv8 model, trained on a dataset enhanced with low-light image processing, achieved notable performance gains, including a 9.8% increase in precision, a 45.7% increase in recall, a 26.8% increase in mAP50, and a 41.0% improvement in mAP50-95, particularly when validated on dark images. These results highlight the effectiveness of incorporating the low light enhancement model into the YOLOv8 training pipeline, enabling the model to better adapt to low-light conditions and produce more accurate pedestrian detection.

Despite these advancements, several limitations indicate opportunities for future research. The separate application of the low-light enhancement model, while beneficial for accuracy, may slow down detection speed. Additionally, the image resizing process risks losing important details, which suggests that more advanced image processing techniques should be explored. Future work should also investigate more sophisticated low-light enhancement models, such as GANs, to further improve detection performance under challenging lighting conditions.

ACKNOWLEDGMENTS

Sincere appreciation to Bina Nusantara University for providing invaluable resources and guidance throughout the research process.

FUNDING INFORMATION

The author would like to express sincere gratitude to the Indonesian Ministry of Education, Culture, Research, and Technology for funding this research under grant number 105/E5/PG.02.00.PL/2024.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Rendi	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			
Devi Fitriana	✓	✓		✓	✓	✓		✓		✓		✓	✓	✓

C : Conceptualization

M : Methodology

So : Software

Va : Validation

Fo : Formal analysis

I : Investigation

R : Resources

D : Data Curation

O : Writing - Original Draft

E : Writing - Review & Editing

Vi : Visualization

Su : Supervision

P : Project administration

Fu : Funding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

INFORMED CONSENT

We have obtained informed consent from all individuals included in this study.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, Rendi, upon reasonable request.

REFERENCES




- [1] A. H. Khan, M. S. Nawaz, and A. Dengel, "Localized Semantic Feature Mixers for Efficient Pedestrian Detection in Autonomous Driving," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2023, pp. 5476–5485, doi: 10.1109/CVPR52729.2023.00530.
- [2] G. L. Hung, M. S. B. Sahimi, H. Samma, T. A. Almohamad, and B. Lahasan, "Faster R-CNN Deep Learning Model for Pedestrian Detection from Drone Images," *SN Computer Science*, vol. 1, no. 2, pp. 1-9, Mar. 2020, doi: 10.1007/s42979-020-00125-y.
- [3] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2016, pp. 779–788, doi: 10.1109/CVPR.2016.91.
- [4] G. Jocher, A. Chaurasia, and J. Qiu, "Ultralytics YOLO," Jan. 2023. [Online]. Available: <https://github.com/ultralytics/ultralytics>. (Accessed: May. 17, 2024).
- [5] H. Zhou, F. Jiang, and H. Lu, "SSDA-YOLO: Semi-supervised domain adaptive YOLO for cross-domain object detection," *Computer Vision and Image Understanding*, vol. 229, Mar. 2023, doi: 10.1016/j.cviu.2023.103649.
- [6] J. Li, Z. Xu, L. Fu, X. Zhou, and H. Yu, "Domain adaptation from daytime to nighttime: A situation-sensitive vehicle detection and traffic flow parameter estimation framework," *Transportation Research Part C: Emerging Technologies*, vol. 124, pp. 1-19, Mar. 2021, doi: 10.1016/j.trc.2020.102946.
- [7] Y. Zheng, D. Huang, S. Liu, and Y. Wang, "Cross-domain Object Detection through Coarse-to-Fine Feature Adaptation," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2020, pp. 13763–13772, doi: 10.1109/CVPR42600.2020.01378.
- [8] Y. Wang *et al.*, "Domain-Specific Suppression for Adaptive Object Detection," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2021, pp. 9598–9607, doi: 10.1109/CVPR46437.2021.00948.
- [9] Y.-J. Li *et al.*, "Cross-Domain Adaptive Teacher for Object Detection," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2022, pp. 7571–7580, doi: 10.1109/CVPR52688.2022.00743.
- [10] X. Kuang *et al.*, "Thermal infrared colorization via conditional generative adversarial network," *Infrared Physics and Technology*, vol. 107, pp. 1-8, Jun. 2020, doi: 10.1016/j.infrared.2020.103338.
- [11] Y. Qiu, Y. Lu, Y. Wang, and H. Jiang, "IDOD-YOLOV7: Image-Dehazing YOLOV7 for Object Detection in Low-Light Foggy Traffic Environments," *Sensors*, vol. 23, no. 3, pp. 1-22, Jan. 2023, doi: 10.3390/s23031347.

Enhancing low-light pedestrian detection: convolutional neural network and YOLOv8 ... (Rendi)




- [12] C. Chen, Q. Chen, J. Xu, and V. Koltun, "Learning to See in the Dark," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, Jun. 2018, pp. 3291–3300, doi: 10.1109/CVPR.2018.00347.
- [13] Z. Fu, Y. Yang, X. Tu, Y. Huang, X. Ding, and K.-K. Ma, "Learning a Simple Low-Light Image Enhancer from Paired Low-Light Instances," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2023, pp. 22252–22261, doi: 10.1109/CVPR52729.2023.02131.
- [14] Y. Han *et al.*, "Low-Illumination Road Image Enhancement by Fusing Retinex Theory and Histogram Equalization," *Electronics (Basel)*, vol. 12, no. 4, pp. 1–18, Feb. 2023, doi: 10.3390/electronics12040990.
- [15] F. Han, J. Yao, H. Zhu, and C. Wang, "Underwater Image Processing and Object Detection Based on Deep CNN Method," *Journal of Sensors*, vol. 2020, pp. 1–20, May 2020, doi: 10.1155/2020/6707328.
- [16] V. Basu, "Low Light Image Enhancement with CNN," Kaggle. [Online]. Available: <https://www.kaggle.com/code/basu369victor/low-light-image-enhancement-with-cnn>. (Accessed: Apr. 23, 2024).
- [17] F. Zhang, Y. Shao, Y. Sun, C. Gao, and N. Sang, "Self-supervised Low-Light Image Enhancement via Histogram Equalization Prior," *Pattern Recognition and Computer Vision (PRCV 2023)*, 2024, pp. 63–75, doi: 10.1007/978-981-99-8552-4_6.
- [18] L. Ma, T. Ma, R. Liu, X. Fan, and Z. Luo, "Toward Fast, Flexible, and Robust Low-Light Image Enhancement," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2022, pp. 5627–5636, doi: 10.1109/CVPR52688.2022.00555.
- [19] S. Ai and J. Kwon, "Extreme Low-Light Image Enhancement for Surveillance Cameras Using Attention U-Net," *Sensors*, vol. 20, no. 2, pp. 1–10, Jan. 2020, doi: 10.3390/s20020495.
- [20] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep Retinex Decomposition for Low-Light Enhancement," *arXiv*, Aug. 2018, doi: 10.48550/arXiv.1808.04560.
- [21] I. Cauldron, "Safer Yolo, in the dark (I)," Medium. [Online]. Available: <https://irenesz.medium.com/safer-yolo-in-the-dark-i-98ddaa7db3ad>. (Accessed: Jun. 06, 2024).
- [22] A. M. Roy and J. Bhaduri, "DenseSPH-YOLOv5: An automated damage detection model based on DenseNet and Swin-Transformer prediction head-enabled YOLOv5 with attention mechanism," *Advanced Engineering Informatics*, vol. 56, Apr. 2023, doi: 10.1016/j.aei.2023.102007.
- [23] A. M. Roy, R. Bose, and J. Bhaduri, "A fast accurate fine-grain object detection model based on YOLOv4 deep neural network," *Neural Computing and Applications*, vol. 34, no. 5, pp. 3895–3921, Mar. 2022, doi: 10.1007/s00521-021-06651-x.
- [24] T.-Y. Lin *et al.*, "Microsoft COCO: Common Objects in Context," *Computer Vision – ECCV 2014 (ECCV 2014)*, 2014, pp. 740–755, doi: 10.1007/978-3-319-10602-1_48.
- [25] J. Delua, "Supervised versus unsupervised learning: What's the difference?," IBM. [Online]. Available: <https://www.ibm.com/think/topics/supervised-vs-unsupervised-learning>. (Accessed: Jul. 06, 2024).
- [26] W. Burger and M. J. Burge, "Histograms and Image Statistics," *Digital Image Processing*, 2022, pp. 29–48, doi: 10.1007/978-3-031-05744-1_2.
- [27] M. Hussain, "YOLOv1 to v8: Unveiling Each Variant—A Comprehensive Review of YOLO," *IEEE Access*, vol. 12, pp. 42816–42833, 2024, doi: 10.1109/ACCESS.2024.3378568.

BIOGRAPHIES OF AUTHORS



Rendy    was born February 10, 2000 in Jakarta, Indonesia. He graduated with a Bachelor's degree in Computer Science from Bina Nusantara University in 2022 and is currently pursuing a Master's degree starting in 2023. Currently he is working as an IT specialist in the banking sector, managing information systems, ensuring data security, and developing technological solutions to enhance banking services. He can be contacted at email: rendi002@binus.ac.id.



Devi Fitrianah    is a lecturer and researcher at the Master of Computer Science Department at Bina Nusantara University. She received her Bachelor's degree in Computer Science from Bina Nusantara University in 2000 and a Master's degree in Information Technology and a Ph.D. degree in Computer Science from Universitas Indonesia in 2008 and 2015 respectively. She has been rewarded with a sandwich program at the Laboratory for Pattern Recognition and Image Processing and GIS (PRIPGIS Lab) Department of Computer Science, Michigan State University, East Lansing, Michigan, USA in 2014. She is now a fellow researcher of the Eureka Robotics Lab, Cardiff Metropolitan University, UK. Her research interests: data mining, machine learning, artificial intelligence, and applied remote sensing. She can be contacted at email: devi.fitrianah@binus.ac.id.